Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 12283–12287, December 1994
Evolution

Dollo's law and the death and resurrection of genes

(development/gene inactivation/functional constraint/evolutionary reversal/molecular evolution)

CHARLES R. MARSHALL*, ELIZABETH C. RAFF, AND RUDOLF A. RAFF

Institute for Molecular and Cellular Biology and Department of Biology, Indiana University, Bloomington, IN 47405

Communicated by Richard E. Dickerson, August 8, 1994 (received for review October 4, 1993)

ABSTRACT **Dollo's law, the concept that evolution is not substantively reversible, implies that the degradation of genetic information is sufficiently fast that genes or developmental pathways released from selective pressure will rapidly become nonfunctional. Using empirical data to assess the rate of loss of coding information in genes for proteins with varying degrees of tolerance to mutational change, we show that, in fact, there is a significant probability over evolutionary time scales of 0.5–6 million years for successful reactivation of silenced genes or "lost" developmental programs. Conversely, the reactivation of long (>10 million years)-unexpressed genes and dormant developmental pathways is not possible unless function is maintained by other selective constraints; the classic example of the resurrection of "hen's teeth" is most likely an experimental artifact, and the experimental reactivation of the *Archaeopteryx* limb developmental program has been shown to be a misinterpretation. For groups undergoing adaptive radiations, lost features may "flicker" on and off, resulting in a distribution of character states that does not reflect the phylogeny of the group.**

As the history of animal life was traced in the fossil record during the 19th century, it was observed that once an anatomical feature was lost in the course of evolution it never staged a return. This observation became canonized as Dollo's law, after its propounder, and is taken as a general statement that evolution is irreversible (reviewed in ref. 1). In 1939, H. J. Muller (2) considered the genetic consequences of evolutionary reversal, and his ideas were extended and united with paleontology by G. G. Simpson in 1953 (3). They argued that since back mutations, recombination, and suppression of mutations all occur, there is no theoretical block to evolutionary reversion. Nonetheless, as in the original statement of Dollo's law, reversals would be statistically improbable if a significant amount of genetic change has accumulated.

Here, we have quantified the probability of evolutionary reversal by estimating the actual time limits over which significant amounts of damage will accumulate in a silenced gene. A silenced gene means a gene from which a transcript is no longer made. Gene silencing does not require deletion of the gene, nor a deleterious mutation in its coding sequences, but can occur through a mutation in its control circuitry that results in loss of expression or expression below a threshold level. Cis-acting regulatory element changes are frequent and result in dramatic modifications of gene expression patterns in various tissues, including loss or gain of expression (4). Particularly good examples are the changes in control circuitry which have occurred in several taxa for genes which encode proteins recruited to serve in the lens (5). Possible mechanisms by which expression might be lost include loss of a binding site for a transcription factor. Function may be regained by back mutation, recombination, or gene conversion. Reactivation may be selected for in some cases, in others merely neutral. If a process or structure is to be regained in evolution by reactivation of a silenced gene, the encoded protein must still be functional. Here, by estimating the rates of inactivation of coding sequences, we estimate the time period over which the coding sequence of a silenced gene could be functionally reactivated. We have taken loss and later regain of expression of a gene as given events and consider how much time can elapse before a protein encoded by a reactivated gene would no longer be functional.

Time Course of Degradation of Coding Information in a Silenced Gene

Studies of pseudogenes show that, in silenced genes, single base changes will predominate early but frameshifts will eventually occur (6–10). We calculated the time course of the degradation of coding information by assuming that the rate of accumulation of nucleotide substitutions in unexpressed genes would occur at rates of neutral substitution and that rates of accumulation of frameshifts would be comparable to observed rates of insertion and deletion in processed pseudogenes.

We have considered changes only in the exons because, for the most part, substitutions, insertions, and deletions in introns are not expected to affect the function of gene. Similarly, we have ignored the degradation of intron/exon splice recognition sites in our calculations because these sites represent a tiny fraction of the total length of the exons of a gene. Moreover, there is a degree of redundancy in splice sites (11).

The probability that a gene retains a functional sequence is given by

$$P_{(\text{retain function})} = P_{(\text{survive point substitutions})} \times P_{(\text{survive frameshifts})}, \quad [1]$$

where $P_{(\text{survive point substitutions})}$ is the probability that a substitution will not result in loss of function (f), raised to the number of substitutions incurred in a given period of time:

$$P_{(\text{survive point substitutions})} = (f)^{r_s l t}, \quad [2]$$

where r_s is the neutral substitution rate [substitutions per site per million years (Myr)], l is the combined length of the gene's exons (bp), and t is time (Myr). $P_{(\text{survive frameshifts})}$ is the probability that a frameshift will not result in loss of function, raised to the number of frameshifts incurred in a given period of time:

$$P_{(\text{survive frameshifts})} = (0.05)^{r_f L t}, \quad [3]$$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbeviation: Myr, million years.

*Present address: Department of Earth and Space Sciences, and Molecular Biology Institute, University of California, Los Angeles, CA 90024-1567.

where r_f is the rate of frameshifts (frameshifts per kb per Myr), L is the exonic length (kb), and t is time (Myr). The value of 0.05 was selected to reflect the conservative assumption that frameshifts in all but the last 5% of a gene are deleterious, based on observations of retention of function of truncated proteins (12, 13). Putting these expressions into Eq. 1, we obtain

$$P_{(\text{retain function})} = (f)^{r_s l t} \times (0.05)^{r_f L t}. \quad [4]$$

Estimating Parameter Values for Eq. 4

Probability That a Random Nucleotide Substitution Is Nonlethal (f). Proteins differ in the stringency of sequence requirements for retention of function; for various proteins, the tolerated number of accepted amino acid replacements per 100 positions per 100 Myr ranges from 0.2 to 143 (14). Different domains within a protein also show different degrees of conservation. For example, in many transcription factors, only the DNA-binding domains are highly conserved. Thus, a mutational study of the *lac* repressor showed that within the DNA-binding domain, 36% of the amino acids tested could not tolerate replacement, whereas outside of this domain, all but 5% of the positions tested could accept replacements (13). However, other critical regulatory proteins, such as the mammalian sex-determining gene *Sry*, exhibit rapid change in the regulatory domains (15).

We directly calculated the value for f for a 17-amino acid region of the λ repressor required for DNA binding but not in contact with DNA, using the data obtained by Bowie *et al.* (16) in an *in vivo* mutational analysis. Within this region, three sites could not be varied without loss of function, nine sites could accept only restricted amino acid replacements, and five sites were unconstrained. By calculating the probability that a random mutation in the coding sequence of this domain (17) would either be synonymous or result in a replacement that the experimental study showed would not cause loss of function, we found that the average value of f over the 17 amino acids was 0.61.

We estimated values of f for a wider range of proteins by using the phylogenetic studies of Fitch and colleagues. Fitch and Markowitz (18) noted that at any given point in evolutionary time a large proportion of codons in cytochrome *c* are not free to vary. The remaining codons are termed concomitantly variable codons, or covarions. The positions of the covarions will change with time, if replacements of covarions constrain positions once free to vary, whereas replacements at other positions may release others from the constraints of complete invariance. Fitch (19) estimated the proportion of covarions (C_{AA}) for several proteins (Table 1).

For each protein in Table 1, we calculated the probability that a random substitution is nonlethal (f) by treating the invariable codons and covarions separately. The value of f for the invariable codons is the proportion of amino acids invariable ($1 - C_{AA}$) times the probability that a random substitution will be synonymous (0.24, if each amino acid codon occurs in a given gene with equal probability and there is no nucleotide composition or substitution bias). As illustrated by the λ repressor domain, covarions include both amino acid positions that are completely unconstrained and positions that are partly constrained. Thus, for the total set of covarions in a protein, the probability that a random substitution is nonlethal equals the proportion of amino acids that are covarions (C_{AA}) times the average probability that a random substitution is nonlethal (p). Combining the invariable and covarion positions, we obtain

$$f = 0.24(1 - C_{AA}) + pC_{AA}. \quad [5]$$

Given the lack of information on the actual functional constraints on covarions (i.e., the value of p), we calculated three values of f for each protein in Table 1. First, as a reasonable value of p for a constrained sequence, we used the value $p = 0.70$ determined from the empirical value of f for the λ repressor domain. Second, to be sure that the values in our model reflect the range in values of p for real proteins, we considered the limiting values of p. An extreme lower limit for p is 0.26, determined from the three positions in the λ repressor domain which were found experimentally to be able to accept only one other amino acid without loss of function (16). As a more realistic lower limit, we used the value of p for the seven most constrained covarions in the λ repressor domain (three sites which can accept only a single replacement and four sites which can accept only a few replacements); in this case p is 0.54. The upper limit for p is the hypothetical case in which all covarions can be replaced with any amino acid without loss of function. For this situation, p is 0.96, reflecting the probability that a substitution will not result in a stop codon. This value remains effectively unchanged even if one takes into account the fact that function may not be lost if a stop codon appears in the last few percent of the gene. We also assumed that there will be no more than one substitution per codon, given that on average only a small number of substitutions over the coding region is usually sufficient for loss of functionality.

Three values of f based on different functional constraints on the covarions (i.e., for values of p of 0.54, 0.70, and 0.96) are given for each protein in Table 1. Since f primarily depends on the value of C_{AA}, these values of p give a reasonable range in values of f, based on available data for

Table 1. Sensitivity of protein coding sequences to nucleotide substitutions

Protein	Length, bp	Fraction of amino acids covarions* (C_{AA})	Probability that random substitution is nonlethal (f)		
			Degree of conservation of covarions (p)		
			Constrained		Unconstrained
			$p = 0.54$	$p = 0.70$	($p = 0.96$)
Hypothetical (all sites vary)	—	1.0	0.54	0.70	0.96
Fibrinopeptide A	57	0.95	0.52	0.68	0.92
λ repressor (partial)	51	0.82	[0.48]	0.61†	[0.83]
Ribonuclease	387	0.60	0.42	0.51	0.67
Insulin C	93	0.58	0.41	0.51	0.66
α Globin	423	0.35	0.35	0.40	0.49
β Globin	438	0.27	0.33	0.36	0.43
Cytochrome *c*	312	0.12	0.27	0.29	0.32
Hypothetical (all invariant)	—	0.0	0.24	0.24	0.24

*Taken from Fitch (19), except for λ repressor (partial), which was calculated from Bowie *et al.* (16).
†This is the empirical value of f for λ repressor (partial), calculated from Bowie *et al.* (16).

real proteins. The minimum possible value of f is 0.24 for a hypothetical protein in which all amino acids must be maintained invariant in order to retain function. Histone H4 approaches this limit, with only four amino acid differences between the human and pea sequences (20). The maximum value of f for a hypothetical protein unconstrained except for maintenance of chain length is 0.96. Given that in all proteins a significant fraction of internal positions must remain hydrophobic to maintain folding, $f = 0.85$ is probably closer to the actual limit (if one-third of the positions require conservation of hydrophobicity, then $p = 0.68$).

Rates of Nucleotide Substitution (r_s) and Frameshifts (r_f). Table 2 gives estimates for neutral rates of nucleotide substitution available for several groups of organisms. Estimates for the rates of frameshifts were based on processed pseudogenes, because these sequences have a high probability of being "dead on arrival." Graur *et al.* (27) list the number of deletions and insertions (n_f) and number of point substitutions per nucleotide site (d) for 22 human, 14 mouse, and 16 rat processed pseudogenes and their functional conspecific homologs. For the human pseudogenes and for the combined mouse and rat data, a correlation was found between n_f/kb and d ($r = 0.60$, and $r = 0.51$, respectively). For the human data the first frameshift occurs at $d = 0.0732$, which, for $r_s = 1 \times 10^{-3}$ substitution per site per Myr for this slowly evolving lineage, corresponds to a waiting time of 73 Myr for the first frameshift/kb to occur, yielding $r_f = 0.014$ frameshift per kb per Myr. For the combined rodent data the first frameshift occurs at $d = 0.0349$, which, for $r_s = 10 \times 10^{-3}$ substitution per site per Myr for the fast evolving lineages, corresponds to a waiting time of 3.5 Myr, yielding $r_f = 0.287$ frameshift per kb per Myr. For lineages evolving at intermediate rates, the value of d was estimated by averaging the d values from the fast and slowly evolving lineages (0.0541), which, for the average value of r_s (5×10^{-3}) gives a waiting time of 10.8 Myr for the first frameshift and $r_f = 0.093$ frameshift per kb per Myr. This ignores the possibility that insertions and deletions may not result in frameshifts, so we slightly underestimate the probability that a gene will remain functional for any given period of time. Note that we have assumed that large insertions and deletions will result in the loss of function even if they do not introduce a frameshift.

Temporal Limits for Reactivation of Silenced Genes. Fig. 1 shows the graphical representation of Eq. 4. Values of f were selected to bracket the realistic range in values for proteins under different selective constraint ($f = 0.7$ and $f = 0.3$). The graphs show that some genes may potentially remain silenced for several million years without mutational inactivation of the encoded protein. This is a surprisingly long time, indicating that reactivation of genes may in fact occur over times that exceed the time required for speciation. The implication for the "direction" of evolution is that even if only a few percent of individuals in a population possess a mutation that

Table 2. Neutral substitution rates for selected groups

Group (ref.)	Neutral substitution rate (r_s), (substitutions per site per Myr) $\times 10^3$
Humans, apes (21)	0.9–1.7
Birds (22, 23)	1–2
Humans, Old World monkeys (21)	1.8–2.8
Goat, pig, cow (21)	2.9–6.0
Higher plants (24)	5.1–7.1
Sea urchins (25)	7.0–8.5
Rodents (21, 26)	3.9–11.8
Mouse, rat (21)	10
Drosophila (26)	12–16

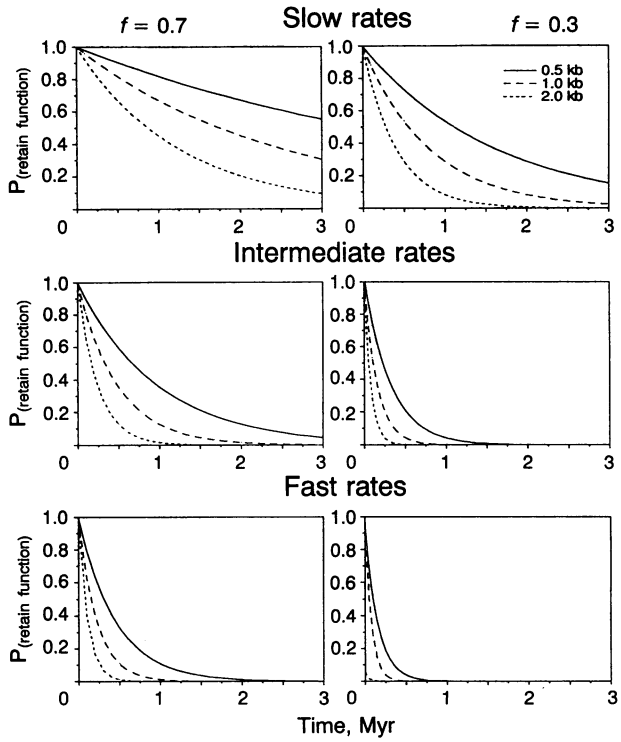

FIG. 1. Probability that reactivated genes will yield functional products ($P_{(\text{retain function})}$) as function of time of absence of selective constraint. Each row represents a calculation based on a different rate of neutral substitution: slow, intermediate, and fast: 1×10^{-3}, 5×10^{-3}, and 10×10^{-3} substitutions per site per Myr, respectively. The lengths of exon modeled are 0.5 kb (——), 1.0 kb (– – –), and 2.0 kb (- - -) (equivalent to 18-kDa, 37-kDa, and 73-kDa proteins).

reactivates a still-viable silenced gene, then resurrection of the gene could occur following a change in expression pattern, given appropriate selective conditions. Population genetics studies show that at only 1% selective advantage, a favored allele goes from a frequency in the population of 0.01 to 0.25 in 350 generations; for a 10% advantage, only 35 generations are required (28).

Intimations of Immortality: Maintenance by Selection

We note that genes no longer expressed in one pathway need not suffer loss of information if they also function in some other pathway in development. Thus there is a plausible mechanism for the cryptic retention of apparently lost genes over longer intervals of time than suggested from rates of accumulation of substitutions and frameshifts in coding sequences. There is a growing set of examples of proteins that serve multiple functions and thus provide the basis for possible evolutionary retention of a protein after loss of one or a subset of its functions. Perhaps the most striking example is the recruitment of metabolic enzymes to serve as structural proteins of the eye lens in both vertebrate and invertebrate groups (5, 29). This phenomenon illustrates the relative ease with which a change in site and timing of expression of a protein may occur, since similar events involving several different proteins have occurred several different times. These proteins function as crystallins in the lens, not as enzymes, yet many of the captured lens proteins still exhibit enzyme activity.

The lens protein story is not a unique case. Other examples of proteins with known multiple functions include the glycosylation signal sequence-binding protein, which is the same protein identified as having at least three other diverse functions in the endoplasmic reticulum lumen (30). The

cation-independent mannose-6-phosphate receptor involved in targeting proteins to lysosomes also functions on the cell surface as a receptor for insulin-like growth factor II (31). A final and significant example is the human KALIG-1 gene, deletion of which causes both sterility and loss of olfaction (32). The gene functions in establishment of two distinct nerve pathways, one in olfactory neurons and the other in neurons that produce the hypothalamic gonadotropin-releasing hormone.

It is also possible for a cryptic, or silenced, gene to be retained without any selective advantage over useless DNA sequences, if selection for the reactivated genes occasionally occurs for a substantially long time (33).

Short-Term Regain of Complex Morphogenetic Pathways

The reactivation of a complex morphogenetic pathway would seem unlikely because the probability of reactivation of multiple genes should be small. Yet, as the examples below illustrate, reactivation of morphogenetic processes occurs, although the genetic mechanisms involved are generally unknown. In the permanently larval salamander, the axolotl, a single gene change has resulted in the loss of production of thyroxine, and thus there is a failure to activate other genes involved in metamorphosis (34). These animals reproduce as overgrown "larvae" (they are neotenic). Yet axolotls undergo metamorphosis when exogenous thyroxine is provided. Thus, the downstream genes for metamorphosis are still functional. There are other permanent neotenes among other groups of salamanders in which metamorphosis has apparently been silenced for a much longer time; these species fail to respond to exogenous thyroxine (35). In Mexico, a complex radiation of ambystomatid salamanders has given rise to metamorphosing, facultative, and permanently neotenic species (36). Although the radiation is species-rich, the overall age of the radiation of these salamanders is only 10–12 Myr. Fig. 2 shows that the nodes between cladistic events are in many cases separated by short branches, which span less than 0.5–1 Myr. Both known reproductive modes and inferred ancestral modes have been placed on the tree; it is clear that within this radiation of salamander species there have been numerous reversals, including both loss and recovery of metamorphosis (36). This complex trait is thus readily reversible over short evolutionary time spans and "flickers" with respect to speciation events.

Is Long-Term Reversal Possible?

Experiments done just over 30 years ago by Hampé (37) have strongly influenced ideas of retention and reactivation of developmental genetic programs in evolution. Modern birds have a greatly reduced fibula that does not extend the full length of the tibia. The Jurassic stem-group ("ancestral") bird *Archaeopteryx* had a much more reptilian leg, with a fibula that articulated at both ends. Hampé embedded a mica flake in a chicken embryo between the regions of the limb bud destined to produce the tibia and that destined to produce the fibula. He concluded that because competition between the two territories was blocked, a full-length fibula was formed with a distal articulation, restoring the ancestral condition. The inference was clear: ancient developmental programs could be retained in silent form for well over 100 Myr. These experiments were repeated and reinterpreted by Müller (38). His results show that what actually occurs is only an apparent lengthening of the fibula due to an experimentally induced shortening of the tibia. No fibular articulation forms. There is no atavistic recovery of the *Archaeopteryx* limb development program.

The last toothed birds date from 70- to 80-Myr-old Cretaceous rocks. Kollar and Fisher (39) asked whether avian pharyngeal epithelium could still produce teeth if provided with inductive signals from mouse oral mesenchyme. Grafts of mouse oral dermis and chick oral epidermis were cultured together. The chick tissue apparently responded by forming enamel organs and even, in a few cases, molariform "teeth." Other investigators have reported conflicting results in obtaining enamel organs or enamel deposition from avian oral epidermis cells cocultured with mouse oral cells (40–42). The interpretation of the recombinant graft experiments between chick and mouse is still clouded by the possibility that the mouse dermis may have been contaminated by mouse epithelial cells. Such hidden contamination has been shown to be a confounding problem in other studies of induction using tissue chimeras (43). The possibility of some response by chick tissue to inductive signals produced by mammalian tissues is consistent with epidermal graft experiments between different classes of vertebrates (44). However, the

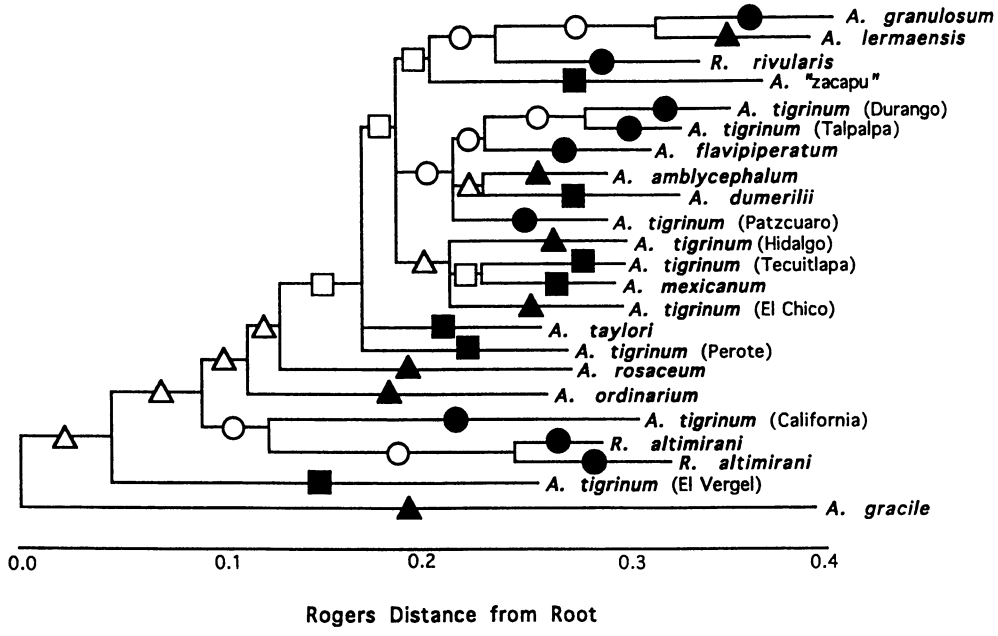

FIG. 2. Flickering of developmental traits during the radiation of Mexican ambystomatid salamanders. Developmental modes of living species are given as solid symbols: circles, metamorphosing; squares, reproduction as neotenic larvae; triangles, facultative. Open symbols represent inferred ancestral developmental modes. The ambystomatid radiation occupied 10–12 Myr ago as inferred from the age of the volcanic uplift that produced the highlands that the salamanders inhabit and from molecular clock estimates. Reversals from larval reproduction to metamorphosis are inferred at two steps in the tree, which occupy spans of <1 Myr. Reproduced from Shaffer (36), with permission (copyright 1984, Society for the Study of Evolution, Lawrence, KS).

enamelin genes would have had to survive in silenced form in birds for some 80 Myr. Although the homolog of the mammalian enamelin gene has been detected in reptiles, it could not be detected in birds (45). There is a near-zero probability of restoring a silent gene over 70–80 Myr, even in lineages with slow evolving genomes (see Fig. 1).

Reversals of long-lost structures do occur but evidently result from the cooption of genes that continue to survive in other roles. Among living cats, *Lynx* possesses primitive field features in its dentition. The carnassial teeth of the lower jaw (M1) in *Lynx* exhibit a third cusp at the hind edge of the tooth. Fossil evidence indicates that this structure was lost and then regained in this lineage (46). *Lynx* also exhibits the reappearance of the second molar, M2. This tooth has been absent in felids since the Miocene, about 20 Myr. Kurtén (46) has suggested that the regain of a cusp on the rear edge of the carnassial and the reappearance of M2 are correlated and result from a "reactivation of the molarization field," which brings M2 above the threshold of realization. Suppression of M2 probably did not involve the loss of any structural genes, since the same genes are presumably required for the morphogenesis of all teeth. In this case, regain of a lost structure after a long period of eclipse may not represent activation of any silenced gene but stem from changes in the level of gene activity controlling the size or strength of the molarization field.

Other reversals have been noted in complex structures, including the reappearance of "lost" muscles in the limbs of some birds (47) and limbs in usually limbless tetrapods (48). It has also been possible to experimentally reverse toe number in guinea pigs by selective breeding from three toes to a more primitive four toes (49, 50). This reversal appears to be the consequence of the continued maintenance of an ancestral developmental pathway that can produce more toes in guinea pigs and can be elicited in the appropriate genetic background. As is the *Lynx* molar, toe number is a meristic trait: once the anlage is provided, the "toe program" is played out automatically (51).

Consequences for Systematics

The methodology of cladistics provides a powerful way of making explicit statements of phylogenetic relationship (52). One of the central tenets of the method is that the evolutionary process produces a hierarchical, or nested, distribution of evolutionary innovations. In practice, determining unambiguous nested sets is difficult, and even the most parsimonious cladograms have considerable character reversals and parallelisms (collectively termed homoplasies). From the cladistic point of view, homoplasy is just noise, and the cladogram that minimizes homoplasy is favored. However, if several speciation events occur within a few million years, it is possible that morphological features may disappear and reappear among the diversifying lineages. Thus, for groups undergoing rapid speciation, morphological innovations will not necessarily produce a simple nested pattern. Flickering has the potential to produce significant homoplasy by parallel reactivation in some lineages.

This work was supported in part by separate grants from the National Institutes of Health to C.R.M., E.C.R., and R.A.R. and by a grant from the National Science Foundation to C.R.M.

1. Gould, S. J. (1970) *J. Hist. Biol.* **3**, 189–212.
2. Muller, H. J. (1939) *Biol. Rev.* **14**, 261–280.
3. Simpson, G. G. (1953) *The Major Features of Evolution* (Columbia Univ. Press, New York), pp. 310–312.
4. Dickinson, J. W. (1988) *BioEssays* **8**, 204–208.
5. Wistow, G. J. (1993) *Trends Biochem. Sci.* **18**, 301–306.
6. Li, W.-H. (1983) in *Evolution of Genes and Proteins*, eds. Rei, M. & Koehn, R. K. (Sinauer, Sunderland, MA), pp. 14–37.
7. Scarpulla, R. C. (1984) *Mol. Cell. Biol.* **4**, 2279–2288.
8. Wu, C. I., Li, W.-H., Shen, J. J., Scarpulla, R. C., Limbach, K. J. & Wu, R. (1986) *J. Mol. Evol.* **23**, 61–75.
9. Qigley, F., Brinkman, H., Martin, W. F. & Cerff, R. (1989) *J. Mol. Evol.* **29**, 412–421.
10. Limbach, K. J. & Wu, R. (1985) *Nucleic Acids Res.* **13**, 631–644.
11. Smith, C. W. J., Patton, J. G. & Nadal-Ginard, B. (1989) *Annu. Rev. Genet.* **23**, 527–577.
12. Fackenthal, J. D., Turner, R. F. & Raff, E. C. (1993) *Dev. Biol.* **158**, 213–227.
13. Miller, J. H., Coulondre, C., Hofer, M., Schmeissner, U., Sommer, H. & Schmitz, A. (1979) *J. Mol. Biol.* **131**, 191–222.
14. Dickerson, R. E. (1971) *J. Mol. Evol.* **1**, 26–45.
15. Whitfield, L. S., Lovell-Badge, R. & Goodfellow, P. N. (1993) *Nature (London)* **364**, 713–715.
16. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990) *Science* **247**, 1306–1310.
17. Sauer, R. T. (1978) *Nature (London)* **276**, 301–302.
18. Fitch, W. M. & Markowitz, E. (1970) *Biochem. Genet.* **4**, 579–593.
19. Fitch, W. M. (1976) *J. Mol. Evol.* **8**, 13–40.
20. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, pp. 49–50.
21. Li, W.-H., Tanimura, M. & Sharp, P. (1987) *J. Mol. Evol.* **25**, 330–342.
22. Britten, R. J. (1986) *Science* **231**, 1393–1398.
23. Krajewski, C. (1990) *Mol. Biol. Evol.* **7**, 65–73.
24. Wolfe, K. H., Sharp, P. M. & Li, W.-H. (1989) *J. Mol. Evol.* **29**, 208–211.
25. Smith, A. B. (1988) *Mol. Biol. Evol.* **5**, 345–365.
26. Sharp, P. M. & Li, W.-H. (1989) *J. Mol. Evol.* **28**, 398–402.
27. Graur, D., Shuli, Y. & Li, W.-H. (1989) *J. Mol. Evol.* **28**, 279–285.
28. Strickberger, M. W. (1976) *Genetics* (Macmillan, New York), 2nd Ed., p. 765.
29. Cuthbertson, R. A., Tomarov, S. I. & Piatigorski, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4004–4008.
30. Geetha-Habib, M., Noiva, R., Kaplan, H. A. & Lennarz, W. J. (1988) *Cell* **54**, 1053–1060.
31. Tong, P. Y., Tollefsen, S. E. & Kornfeld, S. (1988) *J. Biol. Chem.* **263**, 2585–2588.
32. Franco, B., Guioli, S., Pragliola, A., Incerti, B., Bardoni, B., Tonlorenzi, R., Carrozzo, R., Maestrini, E., Pieretti, M., Taillon-Miller, P., Brown, C. J., Willard, H. F., Lawrence, C., Persico, M. G., Camerino, G. & Ballabio, A. (1991) *Nature (London)* **353**, 529–536.
33. Li, W.-H. (1984) *Mol. Biol. Evol.* **1**, 213–219.
34. Tomkins, R. (1978) *Am. Zool.* **18**, 313–319.
35. Dent, J. N. (1968) in *Metamorphosis*, eds. Etkin, W. & Gilbert, L. I. (Appleton-Ashbury-Crofts, New York), pp. 271–311.
36. Shaffer, H. B. (1984) *Evolution* **38**, 1194–1206.
37. Hampé, A. (1960) *J. Embryol. Exp. Morphol.* **8**, 241–245.
38. Müller, G. B. (1989) *J. Evol. Biol.* **2**, 31–47.
39. Kollar, E. J. & Fisher, C. (1980) *Science* **207**, 993–995.
40. Cummings, E. G., Bringas, P., Jr., Grodin, M. S. & Slavkin, H. S. (1981) *Differentiation* **20**, 1–9.
41. Lemus, D., Coloma, L., Fuenzalida, M., Illanes, J., Paz De La Vega, Y., Ondarza, A. & Blanquez, M. J. (1986) *J. Morphol.* **189**, 121–129.
42. Fuenzalida, M., Lemus, R., Romero, S., Fernandez-Valencia, R. & Lemus, D. (1990) *J. Exp. Zool.* **256**, 264–272.
43. Henry, J. J. & Grainger, R. M. (1987) *Dev. Biol.* **124**, 200–214.
44. Dhouailly, D. (1975) *Rouxs Arch. Dev. Biol.* **177**, 323–340.
45. Lyngstadaas, S. P., Risnes, S., Noedbo, H. & Floves, A. G. (1990) *J. Comp. Physiol. B* **160**, 469–472.
46. Kurtén, B. (1963) *Fin. Vetenshaps-Soc. Helsingfors. Commentat. Biol.* **26**, 1–12.
47. Raikow, R. J., Borecky, S. J. & Berman, S. L. (1979) *Condor* **81**, 203–206.
48. Lande, R. (1978) *Evolution* **32**, 73–92.
49. Wright, S. (1934) *Genetics* **19**, 506–536.
50. Wright, S. (1934) *Genetics* **19**, 537–551.
51. Raff, R. A. & Kaufman, T. C. (1983) *Embryos, Genes, and Evolution* (Macmillan, New York).
52. Wiley, E. O. (1981) *Phylogenetics* (Wiley–Liss, New York).